(19) Europäisches Patentamt
European Patent Office
Office européen des brevets

(11) Publication number : **0 360 674 B1**

(12) EUROPEAN PATENT SPECIFICATION

(45) Date of publication of patent specification : **15.11.95 Bulletin 95/46**

(51) Int. Cl.^{6} : **G06F 15/80**

(21) Application number : **89402544.4**

(22) Date of filing : **15.09.89**

(54) Signal processing system and learning processing system.

Divisional application 94107842.0 filed on 15/09/89.

(30) Priority : **17.09.88 JP 232845/88**
17.09.88 JP 232846/88
20.09.88 JP 235441/88

(43) Date of publication of application : **28.03.90 Bulletin 90/13**

(45) Publication of the grant of the patent : **15.11.95 Bulletin 95/46**

(84) Designated Contracting States : **DE FR GB**

(56) References cited :
D.E. RUMELHART et al.: "Parallel distributed processing", vol. 1:"Foundations", 1986, pages 318-362, The MIT Press,Cambridge, London, GB

(56) References cited :
IEEE INTERNATIONAL CONFERENCE ON NEURAL NETWORKS, San Diego, California, 24th -27th July 1988, pages II-235-242, IEEE; M.F. TENORIO et al.: "Adaptive networks as a model for human speech development"
IEEE FIRST INTERNATIONAL CONFERENCE ON NEURAL NETWORKS, San Diego, California, 21st - 24th June 1987, pages II-251/258; F.E. NORROD et al.: "Feedback-induced sequentiality in neural networks"
IEEE INTERNATIONAL CONFERENCE ON NEURAL NETWORKS, San Diego, California, 24th -27th July 1988, pages 225-230; B. MACUKOW et al.: "Neural network model using a normalized inner product as a measure of similarity"

(73) Proprietor : **SONY CORPORATION**
7-35 Kitashinagawa 6-chome
Shinagawa-ku
Tokyo 141 (JP)

(72) Inventor : **Hiraiwa, Atsunobu**
c/o Sony Corporation
7-35 Kitashinagawa 6-chome
Shinagawa-ku Tokyo (JP)

(74) Representative : **Thévenet, Jean-Bruno et al**
Cabinet Beau de Loménie
158, rue de l'Université
F-75340 Paris Cédex 07 (FR)

Description

BACKGROUND OF THE INVENTION

Field of the Invention

This invention relates to a signal processing apparatus or system carrying out signal processing with the use of a so-called neural network made up of a plurality of units each taking charge of signal processing corresponding to that of a neuron, and a learning processing apparatus or system causing a signal processing section by said neural network to undergo a learning processing in accordance with the learning rule of back propagation.

Prior Art

The learning rule of back propagation, which is a learning algorithm of the neural network, has been tentatively applied to signal processing, Including high speed image processing or pattern recognition, as disclosed In "parallel Distributed Processing", vol. 1, The MIT Press, 1986 or "Nikkel Electronics, issue of August 10, 1987, No. 427, pp 115 to 124. The learning rule of back propagation is also applied, as shown In Fig. 1, to a multistorey neural network having an intermediate layer 2 between an inputlayer 1 and an output layer 3.

Each unit u_j of the neural network shown In Fig. 1 issues an output value which is the total sum net_j of output values O_i of units u_i coupled to the unit u_j by coupling coefficients W_{ji}, transformed by a predetermined function f, such as a sigmold function. That is, when the value of a pattern p is supplied as an input value to each unit u_j of the input layer 1, an output value O_{pj} of each unit u_j of the intermediate layer 2 and the output layer 3 is expressed by the following formula (1)

$$O_{pj} = f_j(net_{pj}) = f_j(\sum_i W_{ji} \cdot O_{pji}) \quad \ldots\ldots (1)$$

The output value O_{pj} of the unit u_j of the output layer 3 may be obtained by sequentially computing the output values of the units u_j, each corresponding to a neuron, from the input layer 1 towards the output layer 3.

In accordance with the back-propagation learning algorithm, the processing of learning consisting In modifying the coupling coefficient W_{ji} so as to minimize the total sum E_p of square errors between the actual output value O_{pj} of each unit u_j of the output layer 3 on application of the pattern p and the desirable output value t_{pj}, that is the teacher signal,

$$E_p = \frac{1}{2} \sum_i (t_{pj} - O_{pj})^2 \quad \ldots\ldots (2)$$

is sequentially performed from the output layer 3 towards the input layer 1. By such processing of learning, the output value O_{pj} closest to the value t_{pj} of the teacher signal is output from the unit u_j of the output layer 3.

If the variant ΔW_{ji} of the coupling coefficient W_{ji} which minimizes the total sum E_p of the square errors is set so that

$$\Delta W_{ji} \propto - \partial E_p / \partial W_{ji} \quad (3)$$

the formula (3) may be rewritten to

$$\Delta W_{ji} = \eta \cdot \delta_{pj} O_{pi} \quad (4)$$

as explained in detail in the above reference materials.

In the above formula (4), η stands for the rate of learning, which is a constant, and which may be empirically determined from the number of the units or layers or from the input or output values. δ_{pj} stands for the error proper to the unit u_j.

Therefore, in determining the above variant ΔW_{ji}, it suffices to compute the error δ_{pj} in the reverse direction, or from the output layer towards the input layer of the network.

The error δ_{pj} of the unit u_j of the output layer 1 is given by the formula (5)

$$\delta_{pj} = (t_{pj} - O_{pj}) f'_j(net_j) \quad (5)$$

On the other hand, the error δ_{pj} of the unit u_j of the intermediate layer 2 may be computed by a recurrent function of the following formula (6)

$$\delta_{pj} = f'j(netj) \sum_k \delta_{pk} W_{kj} \qquad \ldots\ldots (6)$$

using the error δ_{pk} and the coupling coefficient W_{kj} of each unit u_k coupled to the unit u_j, herein each unit of the output layer 3. The process of finding the above formulas (5) and (6) is explained in detail in the above reference materials.

In the above formulas, $f'_j(net_j)$ stands for the differentiation of the output function $f_j(net_j)$.

Although the variant ΔW_{ji} may be found from the above formula (4), using the results of the formulas (5) and (6), more stable results may be obtained by finding it from the following formula (7)

$$\Delta W_{ji(n+1)} = \eta \cdot \delta_{pj} O_{pi} + \alpha \cdot \Delta W_{ji(n)} \quad (7)$$

with the use of the results of the preceding learning. In the above formula, α stands for a stabilization factor for reducing the error oscillations and accelerating the convergence thereof.

The above described learning is repeated until it is terminated at the time point when the total sum E_p of the square errors between the output value O_{pj} and the teacher signal t_{pj} becomes sufficiently small.

It is noted that, in the conventional signal processing system in which the aforementioned back-propagation learning rule is applied to the neural network, the learning constant is empirically determined from the numbers of the layers and the units corresponding to neurons or the input and output values, and the learning is carried out at the constant learning rate using the above formula (7). Thus the number of times of repetition n of the learning until the total sum E_p of the square errors between the output value O_{pj} and the teacher signal t_{pj} becomes small enough to terminate the learning may be enormous to render the efficient learning unfeasible.

Also, the above described signal processing system is constructed as a network consisting only of feed-forward couplings between the units corresponding to the neurons, so that, when the features of the input signal pattern are to be extracted by learning the coupling state of the above mentioned network from the input signals and the teacher signal, it is difficult to extract the sequential time series pattern or chronological pattern of the audio signals fluctuating on the time axis.

In addition, while the processing of learning of the above described multistorey neural network in accordance with the back-propagation learning rule has a promisingly high functional ability, it may occur frequently that an optimum global minimum is not reached, but only a local minimum is reached, in the course of the learning process, such that the total sum E_p of the square errors cannot be reduced sufficiently.

Conventionally, when such local minimum is reached, the initial value of the learning rate η is changed and the processing of learning is repeated until finding the optimum global minimum. This results in considerable fluctuations and protractions of the learning processing time.

Objects of the Invention

It is a primary object of the present invention to provide a signal processing system in which the number of times of repetition of learning until termination of learning may be reduced to realize a more efficient learning.

It is a second object of the present invention to provide a signal processing system whereby the features of the sequential time-series patterns of, for example, audio signals, fluctuating on the time axis, may be extracted by learning of the coupling states in a network constituted by plural units corresponding to neurons.

Summary of the Invention

For accomplishing the primary object of the present invention, the present invention provides a signal processing system according to claim 1.

For accomplishing the second object, the present invention provides a signal processing system according to claim 2.

The above and other objects and novel features of the present invention will become apparent from the following detailed description of the invention which is made in conjuction with the accompanying drawings and the new matter pointed out in the claims.

BRIEF DESCRIPTION OF THE DRAWINGS

Fig. 1 is a diagrammatic view showing the general construction of a neural network to which the back-propagation learning rule is applied.

Fig. 2 is a block diagram schematically showing the construction of a signal processing system according to a first embodiment of the present invention.

Fig. 3 is a flow chart showing the process of learning processing in the learning processing section constituting the signal processing system according to the embodiment shown in Fig. 2.

Fig. 4 is a block diagram schematically showing the construction of a signal processing system according to a second embodiment of the present invention.

Fig. 5 is a diagrammatic view of a neural network showing the construction of the signal processing section of the signal processing system according to the embodiment shown in Fig. 4.

Fig. 6 is a flow chart showing the process of learning processing in the learning processing section constituting the signal processing system of the embodiment shown in Fig. 4.

Fig. 7 is a block diagram schematically showing the construction of a learning processing system in which the present invention may be incorporated.

Figs. 8A and 8B are diagrammatic views showing the state of the signal processing section at the start and in the course of learning processing in the learning processing system shown in Fig. 7.

Fig. 9 is a flow chart showing a typical process of learning processing in the learning processing section constituting the learning processing system shown in Fig. 7.

Fig. 10 is a chart showing the typical results of tests of learning processing on the signal processing section of the neural network shown in Fig. 5 by the learning processing section of the learning processing system of Fig. 7.

Fig. 11 is a chart showing the results of tests of learning on the signal processing section of the neural network shown in Fig. 5, with the number of units of the intermediate layer fixed at six.

Fig. 12 is a chart showing the results of tests of learning on the signal processing system of the neural network shown in Fig. 5, with the number of units of the intermediate layer fixed at three.

DETAILED DESCRIPTION OF THE EMBODIMENTS

By referring to the drawings, certain preferred embodiments of the present invention will be explained in more detail.

The signal processing system of the present invention includes, a shown schematically in Fig. 2, a signal processing section 10 for producing an output value O_{pj} from input signal patterns p and a signal processing section 20 for executing learning for producing an output value O_{pj} closest to the desired output value t_{pj} from the input signal patterns p by the signal processing section 10.

The signal processing section 10 is formed by a neural network including at least an input layer L_I, an intermediate layer L_H and an output layer L_O. These layers L_I, L_H and L_O are made up of units u_{I1} to u_{Ix}, u_{H1} to u_{Hy} and u_{O1} to u_{Oz}, each corresponding to a neuron, wherein x, y and z each represent an arbitrary number.

Each of the units u_{I1} to u_{Ix}, u_{H1} to u_{Hy} and u_{O1} to u_{Oz} is designed to issue an output O_{pj} represented by a sigmoid function according to the formula (8)

$$O_{pj} = \frac{1}{1 + e^{-(net_j + \theta_j)}} \qquad (8)$$

for the total sum net_j of inputs represented by the formula (9)

$$net_j = \sum_i W_{ji} O_{pi} \qquad \ldots\ldots (9)$$

where θ_j stands for a threshold value.

The learning processing section 20 is fed with a desired output value t_{pj} as a teacher signal for the output value O_{oj} of the output layer L_O for the input signal patterns p entered into the signal processing section 10. This learning processing section 20 causes the signal processing section 10 to undergo learning processing of the coupling coefficient W_{ji}, in such a manner that, according to the sequence of steps shown by the flow chart of Fig. 3, the coefficient W_{ji} of the coupling strength between the units u_{I1} to u_{Ix}, u_{H1} to u_{Hy} and u_{O1} to u_{Oz} is sequentially and repeatedly computed from the output layer L_O towards the input layer L_I, until the sum of the quadratic errors between the desired output value t_{pj} and the actual output value O_{oj} become sufficiently small, in order that the output value O_{oj} of the output layer L_O will be closest to the desired output value t_{pj} sup-

plied as the teacher signal.

Thus, in step 1, the learning processing section 20 affords the coupling coefficient W_{ji} to each of the units u_{H1} to u_{Hy} and u_{O1} to u_{Oz} to compute the output value O_{oj} of the output layer L_O for the input signal patterns p in the signal processing section 10. In step 2, the section 20 executes decision as to the converging condition for the actual output value O_{oj}, on the basis of the total sum E_p of the square errors between the actual output value O_{oj} and the desired output value t_{pj} supplied as the teacher signal.

In the decision step 2, it is decided whether the output value O_{oj} obtained at the output layer L_O of the signal processing section 10 is closest to the desired output value t_{pj}. If the result of decision at step 2 is YES, that is, when the total sum E_p of the square errors becomes sufficiently small and the output value O_{oj} is closest to the desired output value t_{pj}, the processing of learning is terminated. If the result of decision is NO, the computation operations of steps 3 through 6 are executed sequentially.

In the next computing step 3, the error δ_{pj} at each of the units u_{H1} to u_{Hy} and u_{O1} to u_{Oz} of the signal processing section 10 is computed. In the computing operation of step 3, the error δ_{oj} of each of the units u_{O1} to u_{Oz} of the output layer L_O is given by the following formula (10):

$$\delta_{oj} = (t_{pj} - O_{oj})O_{oj}(1 - O_{oj}) \quad (10)$$

On the other hand, the error δ_{Hj} of each of the units u_{H1} to u_{Hy} of the intermediate layer L_H is given by the following formula (11):

$$\delta_{Hj} = O_{Hj}(1 - O_{Hj}) \sum_k \delta_{ok} \cdot W_{kj} \quad \ldots\ldots \quad (11)$$

In the next computing step 4, the learning variable β_j of the coefficient W_{ji} of the coupling strength from the i'th one to the j'th one of the units u_{H1} to u_{Hy} and u_{O1} to u_{Oz} is computed as a reciprocal of the square sum of the totality of the inputs added to by 1 as the threshold value, that is, in accordance with the following formula (12):

$$\beta_j = \frac{1}{\sum_i O_{pi}^2 + 1} \quad \ldots\ldots \quad (12)$$

Then, in the computing step 5, the variant ΔW_{ji} of the coupling coefficient W_{ji} from the i'th one to the j'th one of the units u_{H1} to u_{Hy} and u_{O1} to u_{Oz} is computed, using the above learning variable β_j, in accordance with the following formula (13)

$$\Delta W_{ji(n+1)} = \eta \cdot \beta(\delta_{pj} O_{pi}) + \alpha \cdot \Delta W_{ji(n)} \quad (13)$$

where η stands for the learning constant and α the stabilization constant for reducing the error oscillations and accelerating the convergence thereof.

Then, in the computing step 6, the coupling coefficient W_{ji} of the units u_{H1} to u_{Hy} and u_{O1} to u_{Oz} is modified, on the basis of the variant ΔW_{ji} of the coupling coefficient W_{ji} computed at step 5, in accordance with the following formula (14):

$$W_{ji} = W_{ji} + \Delta W_{ji} \quad (14)$$

Then, revert to step 1, the output value O_{oj} of the output layer L_O for the input patterns p at the signal processing section 10 is computed.

The learning processing section 20 executes the above steps 1 through 6 repeatedly, until the learning processing is terminated by the decision at step 2 when the total sum E_p of the square error between the desired output t_{pj} afforded as the teacher signal and the output value O_{oj} becomes sufficiently small and the output value O_{oj} obtained at the output layer L_O of the signal processing section 10 is closest to the desired output value t_{pj}.

In this manner, in the signal processing system of the present first embodiment, the learning constant η is normalized by the above learning variable β represented by the reciprocal of the square sum of the input value O_{pi} at each of the units u_{H1} to u_{Hy} and u_{O1} to u_{Oz} added to by 1 as the threshold value. This causes the learning rate to be changed dynamically as a function of the input value O_{pi}. By performing the learning processing of the coupling coefficient W_{ji} with the learning rate changed dynamically in this manner as a function of the input value O_{pi}, it becomes possible to reduce the number n of times of learning significantly to one fourth to one tenth of that in the case of the conventional learning processing.

It is noted that, by representing the learning constant η and the stabilizing constant α in the formula 13 as the function of the maximum error E_{max} for the input patterns as a whole, as shown by the formulas (15) and

(16):

$$\eta = a E_{max} \quad (15)$$

$$\alpha = -b E_{max} + c \quad (16)$$

where a, b and c are constants, and by changing them dynamically, it becomes possible to perform faster learning processing.

According to the above described first embodiment of the signal processing system, the learning constant η is normalized by the learning variable β represented by the reciprocal of a square sum of the actual input O_{pi} in each unit added to by 1 as a threshold value to cause the learning rate to be changed dynamically in accordance with the input value O_{pi} to execute the learning processing of the coupling coefficient W_{ji} so that it becomes possible to perform stable and fast learning.

A second illustrative of the signal processing system according to the present invention will be hereinafter explained.

As shown schematically in Fig. 4, the signal processing system of the present illustrative embodiment includes a signal processing section 30 for obtaining the output value O_{pj} from the input signal patterns p and a learning processing section 40 for causing the signal processing section 30 to undergo learning to obtain the output value O_{pj} closest to the desired output value t_{pj} from the input signal patterns p.

The signal processing section 30 is formed, as shown in Fig. 5, by a neural network of a three-layer structure including at least an input layer L_I, an intermediate layer L_H and an output layer L_O. These layers L_I, L_H and L_O are constituted by units u_{I1} to u_{Ix}, u_{H1} to u_{Hy} and u_{O1} to u_{Oz}, each corresponding to a neuron, respectively, where x, y and z stand for arbitrary numbers. Each of the units u_{H1} to u_{Hy} and u_{O1} to u_{Oz} of the intermediate layer L_H and the output layer L_O is provided with delay means and forms a recurrent network including a loop LP having its output $O_{j(t)}$ as its own input by way of the delay means and a feedback FB having its output $O_{j(t)}$ as an input to another unit.

In the signal processing system 30, with the input signal patterns p entered into each of the units u_{I1} to u_{Ix} of the input layer L_I, the total sum net_j of the inputs to the units u_{H1} to u_{Hy} of the intermediate layer L_H is given by the following formula (17):

$$\begin{aligned} net_j = & \sum_{e}^{x} \sum_{k=0}^{NI} W_{jx*k+e}\, O_{ie(t-k)} \\ & + \sum_{e}^{y} \sum_{k=1}^{NH} W_{jy*k+i}\, O_{hi(t-k)} \\ & + \sum_{i}^{z} \sum_{k=i}^{NO} W_{jz*k+i}\, O_{oi(t-k)} \\ & + \theta_j \qquad \ldots\ldots (17) \end{aligned}$$

Each of the units u_{H1} to u_{Hy} of the intermediate layer L_H issues, for the total sum net_j of the input signals, an output value $O_{Hj(t)}$ represented by the sigmoid function of the following formula (18):

$$O_{Hj(t)} = \frac{1}{1 + e^{-net_j}} \quad (18)$$

The total sum net_j of the inputs to the units u_{O1} to u_{Oz} of the output layer L_O is given by the following formula (19):

$$net_j = \sum_{i}^{x} \sum_{k=0}^{NH} W_{jx*k+i}\, O_{Hi(t-k)} + \sum_{i}^{z} \sum_{k=1}^{NO} W_{jz*k+i}\, O_{Hi(t-k)} + \theta_j \quad \ldots\ldots (19)$$

while each of the units u_{O1} to u_{Oz} of the output layer L_O issues, for the total sum net_j of the inputs, an output value $O_{oj(t)}$ represented by the following formula (20):

$$O_{oj(t)} = \frac{1}{1 + e^{-net_j}} \quad (20)$$

where θ_j stands for a threshold value and NI, NH and NO stand for the numbers of the delay means provided in the layers L_I, L_H and L_O, respectively.

The learning processing section 40 computes the coefficient W_{ji} of coupling strength between the units u_{O1} to U_{Oz}, u_{H1} to u_{Hy} and u_{I1} to u_{Ix}, from the output layer L_O towards the input layer L_I, sequentially and repeatedly, according to the sequence shown in the flow chart of Fig. 6, while executing the learning processing of the coupling coefficient W_{ji} so that the total sum of the square errors LMS between the desired output value t_{pj} afforded as the teacher signal and the output value O_{oj} of the output layer L_O will be sufficiently small. By such learning processing, the learning processing section 40 causes the output value O_{oj} of the output layer L_O to be closest to the desired output value t_{zr}, afforded as the teacher signal patterns, for an input signal pattern $p_{(xr)}$ supplied to the signal processing section 30. This pattern $p_{(xr)}$ represents an information unit as a whole which fluctuates along the time axis and represented by the xr number of data, where r stands for the number of times of sampling of the information unit and x the number data in each sample.

That is, the section 40 affords at step 1 the input signal patterns $p_{(xr)}$ to each of the units u_{I1} to u_{Ix} of the input layer L_I, and proceeds to computing at step 2 each output value $O_{pj(t)}$ of each of the units u_{H1} to u_{Hy} and u_{O1} to u_{0z} of the intermediate layer L_H and the output layer L_O.

The section 40 then proceeds to computing at step 3 the error δ_{pj} of each of the units u_{O1} to u_{Oz} and u_{H1} to U_{Hy}, from the output layer L_O towards the input layer L_I, on the basis of the output values $O_{pj(t)}$ and the desired output value t_{zr} afforded as the teacher signal.

In the computing step 3, the error δ_{oj} of each of the units u_{O1} to u_{Oz} of the output layer L_O is given by the following formula (21):

$$\delta_{oj} = (t_{pj} - O_{oj})O_{oj}(1 - O_{oj}) \quad (21)$$

wherein the error δ_{pj} of each of the units u_{H1} to u_{Hy} of the intermediate layer L_H is given by the following formula (22):

$$\delta_{Hj} = O_{Hj}(1 - O_{Hj})\sum_{k} \delta_{ok} W_{kj} \quad \ldots\ldots (22)$$

Then, in step 4, the learning variable β_j of the coefficient W_{ji} of coupling strength from the i'th one to the j'th one of the units u_{I1} to u_{Ix}, u_{H1} to u_{Hy} and u_{O1} to u_{Oz} is computed by the following formula (23)

$$\beta_j = \frac{1}{\sum_{i} O_{pi}^{2} + 1} \quad \ldots\ldots (23)$$

in which the learning variable β_j is represented by the reciprocal of the square sum of the input values added to by 1 as a threshold value.

Then in step 5, using the learning variable β_j computed in step 4, the variant ΔW_{ji} of the coupling coefficient w_{ji} from the i'th one to the j'th one of the units u_{O1} to u_{0z}, u_{H1} to u_{Hy} and u_{I1} to u_{Ix} is computed in accordance with the following formula (24) :

$$\Delta w_{ji(n)} = \eta \cdot \beta(\delta_{pj} O_{pi}). \quad (24)$$

In the formula, η stands for a learning constant.

Then, in step 5, the total sum LMS of the square errors of the units with respect to the teacher signal is computed in accordance with the formula (25)

$$LMS = \sum_{p=1} \sum_{i=1} (t_{pi} - O_{pi}) \qquad (25).$$

Then, in step 6, it is decided whether the processing of the steps 1 through 5 has been performed on the R-number of input signal patterns p_{xr}. If the result of decision at step 6 is NO, the section 40 reverts to step 1. When the result of decision at step 6 is YES, that is, when all of the variants ΔW_{ji} of the coupling coefficient W_{ji} between the units u_{O1} to u_{Oz}, u_{H1} to u_{Hy} and u_{I1} to u_{Ix} are computed for the input signal patterns p_{xr}, the section 40 proceeds to step 7 to execute decision of the converging condition for the output value O_{oj} obtained at the output layer L_O on the basis of the total sum LMS of square errors between the output value O_{oj} and the desired output value t_{pj} afforded as the teacher signal.

In the decision step 7, it is decided whether the output value O_{oj} obtained at the output layer L_O of the signal processing section 30 is closest to the desired output value t_{pj} afforded as the teacher signal. When the result of decision at step 7 is YES, that is, when the total sum LMS of the square errors is sufficiently small and the output value O_{oj} is closest to the desired output value t_{pj}, the learning processing is terminated. If the result of decision at step 7 is NO, the section 40 proceeds to computing at step 8.

In this computing step 8, the coupling coefficient W_{ji} between the units u_{O1} to u_{Oz}, u_{H1} to u_{Hy} and u_{I1} to u_{Ix} is modified, on the basis of the variant ΔW_{ji} of the coupling coefficient W_{ji} computed at step 5, in accordance with the following formula (26)

$$\Delta W_{ji(n)} = \Delta W_{ji(n)} + \alpha \Delta W_{ji(n-1)} \quad (26)$$

and the following formula (27)

$$W_{ji(n+1)} = W_{ji(n)} + \Delta W_{ji(n)} \quad (27).$$

After the computing step 8, the section 40 reverts to step 1 to execute the operation of steps 1 to 6.

Thus the section 40 executes the operations of the steps 1 to 8 repeatedly and, when the total sum LMS of the square errors between the desired output value t_{pj} and the actual output value O_{oj} becomes sufficiently small and the output value O_{oj} obtained at the output layer L_O of the signal processing section 30 is closest to the desired output value t_{pj} afforded as the teacher signal, terminates the processing of learning by the decision at step 7.

In this manner, in the present second embodiment of the signal processing system, the learning as to the coupling coefficient W_{ji} between the units u_{O1} to U_{Oz}, u_{H1} to U_{Hy} and u_{I1} to u_{Ix} of the signal processing section 30 constituting the recurrent network inclusive of the above mentioned loop LP and the feedback FB is executed by the learning processing section 40 on the basis of the desired output value t_{pj} afforded as the teacher signal. Hence, the features of the sequential time-base input signal pattern p_{xr}, such as audio signals, fluctuating along the time axis, may also be extracted reliably by the learning processing by the learning processing section 40. Thus, by setting the coupling state between the units u_{O1} to u_{Oz}, u_{H1} to u_{Hy} and u_{I1} to u_{Ix} of the signal processing section 30 by the coupling coefficient W_{ji}, obtained as the result of learning by the learning processing section 40, the time-series input signal pattern p_{xr} can be subjected to desired signal processing by the signal processing section 30.

Moreover, in the second illustrative embodiment of the present invention, similarly to the previously described first embodiment, the learning constant η is normalized by the learning variable β indicated as the reciprocal of the square sum of the input values at the units u_{H1} to u_{Hy} and u_{O1} to u_{Oz}, and the learning processing as to the coupling coefficient W_{ji} is performed at the dynamically changing learning rate, as a function of the input value O_{pi}, so that learning can be performed stably and expeditiously with a small number of times of learning.

In this manner, in the present second embodiment of the signal processing system, signal processing for input signals is performed at the signal processing section 30 in which the recurrent network inclusive of the loop LP and the feedback FB is constituted by the units u_{H1} to u_{Hy} and u_{O1} to u_{Oz} of the intermediate layer L_H and the output layer L_O each provided with delay means. In the learning processing section 40, the learning as to the coupling state of the recurrent network by the units u_{H1} to u_{Hy} and u_{O1} to u_{Oz} constituting the signal processing section 30 is executed on the basis of the teacher signal. Thus the features of the sequential time-base patterns, fluctuating along the time axis, such as audio signals, can be extracted by the above mentioned learning processing section to subject the signal processing section to the desired signal processing.

A preferred illustrative embodiment learning processing system in which the present invention can be incorporated will be hereinafter explained.

The basic construction of this learning processing system is shown in Fig. 7. As show therein, the system includes a signal processing section 50 constituted by a neural network of a three-layered structure including at least an input layer L_I, an intermediate layer L_H and an output layer L_O, each made up of plural units performing a signal processing corresponding to one of a neuron, and a learning processing section 60 subjecting the learning processing to the signal processing consisting in sequentially repeatedly computing the coefficient W_{ji} of coupling strength between the above units from the output layer L_O towards the input layer L_I on the basis of the error data δ_{pj} between the output value of the output layer L_O and the desired output value O_{pj} afforded as the teacher signal t_{pj}, for the input signal patterns p entered into the input layer L_I of the signal processing section 50, and learning the coupling coefficient W_{ji} in accordance with the back-propagation learning rule.

The learning processing section 60 executes the learning processing of the coupling coefficient W_{ji} as it causes the number of the units of the intermediate layer L_H of the signal processing section 50 to be increased, and thus the section 60 has the control function of causing the number of units of the intermediate layer L_H to be increased in the course of learning processing of the coupling coefficient W_{ji} The learning processing section 60 subjects the signal processing section 50 having the input layer L_I, an intermediate layer L_H and an output layer L_O made up of arbitrary numbers x, y and z of units u_{I1} to u_{Ix}, u_{H1} to u_{Hy} and u_{O1} to u_{Oz}, each corresponding to a neuron, respectively, as shown in Fig. 8A, to learning processing as to the coupling coefficient W_{ji}, while the section 60 causes the number of the units L_H in the intermediate layer to be increased sequentially from y to (y+m), as shown in Fig. 8B.

It is noted that the control operation of increasing the number of the units of the intermediate layer L_H may be performed periodically in the course of learning processing of the coupling coefficient W_{ji}, or each time the occurrence of the above mentioned local minimum state is sensed.

The above mentioned learning processing section 60, having the control function of increasing the number of the units of the intermediate layer L_H in the course of learning processing of the coupling coefficient W_{ji}, subjects the signal processing section 50 formed by a neural network of a three-layer structure including the input layer L_I, intermediate layer L_H and the output layer L_O to the learning processing of the coupling coefficient W_{ji}, as it causes the number of units of the intermediate layer L_H to be increased. Thus, even on occurrence of the local minimum state in the course of learning of the coupling coefficient W_{ji}, the section 50 is able to increase the number of the units of the intermediate layer L_H to exit from such local minimum state to effect rapid and reliable convergence into the optimum global minimum state.

Tests were conducted repeatedly, in each of which the learning processing section 50 having the control function of increasing the number of units of the intermediate layer in the course of leaming of the coupling coefficient W_{ji} causes the signal processing section 60 constituting the recurrent network including the feedback FB and the loop LP in the second embodiment of the signal processing system to undergo the process of learning the coefficient W_{ji}, with the number of the units of the input layer L_I of 8(x=8), that of the output layer L_O of 3(z=3), the number of the delay means of each layer of 2 and with the input signal pattern p during learning operation, using 21 time-space patterns of 1=8x7, and the processing algorithm shown in the flow chart of Fig. 9, with the learning being started at the number of the units of the intermediate layer L_H of 3(y=3) and with the number of the units of the Intermediate layer L_H being increased during the learning process. By increasing the number of the units of the intermediate layer L_H three to five times, the test results were obtained in which the convergence to the optimum global minimum state were realized without going into the local minimum state.

Fig. 10 shows, as an example of the above tests, the test results in which learning processing of converging into the optimum minimum state could be achieved by adding the units of the intermediate layer L_H at the timing shown by the arrow mark in the figure and by increasing the number of units of the intermediate layer L_H from three to six. The ordinate in Fig. 10 stands for the total sum LMS of the quadratic errors and the abscissa stands for the number of times of the learning processing operations.

The processing algorithm shown in the flow chart of Fig. 9 is explained.

In this processing algorithm, in step 1, the variable K indicating the number of times of the processing for detecting the local minimum state is initialized to "0", while the first variable Lms for deciding the converging condition of the learning processing is also initialized to 10000000000.

Then, in step 2, the variable n indicating the number of times of learning of the overall learning pattern, that is, the l-number of the input signal patterns p, is initialized. The program then proceeds to step 3 to execute the learning processing of the l-number of the input signal patterns p.

Then, in step 4, decision is made of the variable n indicating the number of times of learning. Unless n=3, the program proceeds to step 5 to add one to n ($n \rightarrow n+1$), and then reverts to step 3 to repeat the learning processing. When n=3, the program proceeds to step 6.

In step 6, after the value of the first variable Lms is maintained as the value of the second variable Lms(-

1) for deciding the converging condition of the learning processing, the total sum of the square errors between the output signal and the teacher signal in each unit is computed in accordance with the formula (28), this value being then used as the new value for the first variable Lms, such that

$$Lms = \sum_{n=1} \sum_{i=1}^{m} (t_{pi} - o_{pi})^2 \qquad \ldots\ldots (28).$$

Then, in step 7, the first variable Lms for deciding the converging condition of the learning processing is compared with the second variable Lms(-1). If the value of the first variable Lms is lesser than that of the second variable Lms(-1), the program proceeds to step 8 to decide whether or not the variable K indicating the number of times of the processing operations for detecting the local minimum state is equal to 0.

If, in step 8, the variable K is 0, the program reverts directly to step 2. If the variable K is not 0, setting of K K+1 is made in step 9. The program then reverts to step 2 to initialize n to 0(n=0) to execute the learning processing of the l-number of the input signal patterns p in step 3.

If, in step 7, the value of the first variable Lms is larger than that of the second variable Lms(-1), the program proceeds to step 10 to set the value of K indicating the number of times of the processing operations for detecting the local minimum state (K → K+1). Then, in step 11, it is decided whether or not the value of K is 2.

If, in step 11, the value of the variable K is not 2, the program reverts directly to step 2. If the variable K is 2, it is decided that the local minimum state is prevailing. Thus, in step 12, control is made for increasing the number of the units of the intermediate layer L_H. Then, in step 13, setting of K=0 is made. The program then reverts to step 2 for setting of n=0 and then proceeds to step 3 to execute the learning processing of the above mentioned l-number of the input signal patterns p.

Test on the learning processing was conducted of the signal processing section 50 of the above described second embodiment of the signal processing system constituting the recurrent network including the feedback loop FB and the loop LP shown in Fig. 5, with the number of the units of the intermediate layer L_H being set to six (y=6). The test results have revealed that the learning processing need be repeated an extremely large number of times with considerable time expenditure until the convergence to the optimum minimum state was achieved, and that the local minimum state prevailed for three out of eight learning processing tests without convergence to the optimum global minimum state.

Fig. 11 shows, by way of an example, the results of the learning processing tests in which the local minimum state was reached.

In this figure, the ordinate stands for the total sum LMS of the square errors and the abscissa stands the number of times of the learning processing operations.

Also the tests on the learning processing was conducted 30 times on the signal processing section 50 of the above described second embodiment of the signal processing system constituting the recurrent network including the feedback loop FB and the loop LP shown in Fig. 5, with the number of the units of the intermediate layer L_H being set to three (y=3). It was found that, as shown for example in Fig. 12, the local minimum state was reached in all of the tests on learning processing without convergence to the optimum global minimum state.

In Fig. 12, the ordinate stands for the total sum LMS of the square errors and the abscissa stands the number of times of the learning processing operations.

From the foregoing, it is seen that the present invention can be incorporated in a learning processing system in which the learning processing of the coefficient of coupling strength is performed, while the number of the units of the intermediate layer is increased by the learning processing section, whereby the convergence to the optimum global minimum state is achieved promptly and reliably to achieve the stable learning processing to avoid the local minimum state in the learning processing process conforming to the back-propagation learning rule.

Claims

1. A signal processing system comprising :

a signal processing section (10; 30) composed of a multi-layer neural network with three layers : an input layer L_I, an intermediate layer L_H, and an output layer L_O, the layers being made up of units u_{I1} to U_{Ix}, u_{H1} to u_{Hy}, u_{O1} to u_{Oz}, respectively, each unit corresponding to a neuron, the network consisting of feed-forward couplings between the units, each of the units j in the intermediate layer and in the output

layer being designed to issue for an input pattern p entered into the input layer an output signal O_{pj} represented by a sigmoid function according to the formula :

$$O_{pj} = 1/(1 + \exp\{-(net_j + \theta_j)\})$$

for the total sum net_j of inputs, where

θ_j is a threshold value,

net_j is the total sum of the inputs to a unit j in the intermediate layer and in the output layer, and O_{pj} is the jth element of the actual output pattern produced by the representation of the input pattern,

the system further comprising

a learning process section (20; 40) executing a learning process using a back-propagating learning algorithm, the process consisting in modifying the coupling coefficients W_{ji} of all units j in the intermediate layer and in the output layer with a variant DELTA W_{ji} so as to minimize the total sum of square errors between the actual output value O_{oj} of unit j in layer L_O produced from an input signal pattern and the desirable output value t_{pj} (teacher signal) for said unit j in the layer L_O, whereby W_{ji} is the weight for the signal from the ith to the jth unit,

the learning process section (20; 40) being fed with a desired output value t_{pj} as a teacher signal for the output value O_{oj} of a unit j in the output layer L_O for the input signal patterns p entered into the input layer,

the learning process section computing the error value for each unit in the output layer and in the intermediate layer,

the error δ_{oj} of each unit j of the output layer L_O being computed by the formula :

$$\delta_{oj} = (t_{pj} - O_{oj})\, O_{oj}\, (1 - O_{oj})$$

the error δ_{Hj} of each unit j of the intermediate layer L_H being computed by the formula :

$$\delta_{Hj} = O_{Hj}(1 - O_{Hj})\, \Sigma_k(\delta_{Ok} \cdot W_{kj})$$

the coupling coefficient W_{ji} of the units in the intermediate layer and in the output layer being given by the formula :

$$W_{ji(n+1)} = W_{ji(n)} + \text{DELTA}\, W_{ji(n)}$$

the learning process being executed repeatedly until the total sum E of the square error between the desired output afforded as the teacher signal and the output signal becomes sufficiently small,

characterized in that :

the learning process section (20; 40) computes a learning variable β_j for each coupling coefficient W_{ji} of each unit j in the intermediate layer and in the output layer for all of its inputs O_i :

$$\beta_j = 1 / (\Sigma(O_{pi}^2) + 1)$$

and in that for all these units in the intermediate layer and in the output layer the variant DELTA W_{ji} of the coupling coefficient W_{ji} is computed by using said learning variable β_j :

$$\text{DELTA}\, W_{ji(n)} = N \cdot \beta_j(\delta_{pj} \,.\, O_{pi}) + \alpha \cdot \text{DELTA}\, W_{ji(n-1)}$$

where N is the learning constant, α is the stabilization constant for reducing the error oscillations, and n is the number of times of learning.

2. The signal processing system according to claim 1, wherein each of the units in the intermediate layer and in the output layer are also provided with further couplings each provided with delay means so forming a recurrent network including :

a loop LP to provide via said delay means its output O_j as one of its inputs, and

a feedback path FB to provide its output O_j as an input to another unit in the same layer or in the intermediate layer.

3. The signal processing system according to claim 1 or 2, wherein control means are provided in said learning processing section (20; 40) for increasing the number of the units of said intermediate layer, and wherein said learning processing section performs learning processing of the coefficient of coupling strength W_{ji} as said learning processing section causes the number of the units of the intermediate layer to be increased.

Patentansprüche

1. Signalverarbeitungssystem

mit einem Signalverarbeitungsabschnitt (10; 30) bestehend aus einem mehrschichtigen neuronalen Netzwerk mit drei Schichten: einer Eingangsschicht L_I, einer Zwischenschicht L_H und einer Ausgangsschicht L_O, wobei diese Schichten aus Einheiten u_{I1} bis u_{Ix}, u_{H1} bis u_{Hy} bzw. u_{O1} bis u_{Oz} bestehen,

wobei jede Einheit einem Neuron entspricht, das Netzwerk aus Vorwärtskoppelverbindungen zwischen den Einheiten besteht, jede der Einheiten j in der Zwischenschicht und in der Ausgangsschicht so ausgebildet ist, daß für ein in die Eingangschicht eingegebenes Eingangsmuster p ein Ausgangswert O_{pj} ausgegeben wird, der durch eine Sigmoid-Funktion (S-Funktion) nach der Formel

$$O_{pj} = 1/(1 + \exp\{-(net_j + \theta_j)\})$$

für die Gesamtsumme net_j von Eingangswerten dargestellt wird, worin

θ_j ein Schwellwert,

net_j die Gesamtsumme der Eingangswerte für eine Einheit j in der Zwischenschicht und in der Ausgangsschicht und

O_{pj} das j-te Element des durch die Darstellung des Eingangsmusters erzeugten tatsächlichen Ausgangsmusters bedeuten,

wobei das System weiterhin aufweist:

einen Lernprozeßabschnitt (20; 40), der unter Benutzung eines sich rückwärts ausbreitenden Lernalgorithmus einen Lernprozeß ausführt, der darin besteht, daß die Kopplungskoeffizienten W_{ji} aller Einheiten j in der Zwischenschicht und in der Ausgangsschicht mit einer Varianten DELTA W_{ji} derart modifiziert werden, daß die Gesamtsumme der quadratischen Fehler zwischen dem von einem Eingangssignalmuster erzeugten tatsächlichen Ausgangswert O_{oj} in der Schicht L_O und dem gewünschten Ausgangswert t_{pj} (Lehrsignal) für die genannten Einheit j in der Schicht L_O zu einem Minimum wird, wobei W_{ji} die Gewichtung für das Signal aus der i-ten Schicht zu der j-ten Schicht bedeutet,

. wobei

der Lernprozeßabschnitt (20; 40) mit einem gewünschten Ausgangswert t_{pj} als Lehrsignal für den Ausgangswert O_{oj} einer Einheit j in der Ausgangsschicht L_O für die in die Eingangschicht eingegebenen Eingangssignalmuster p gespeist wird,

der Lernprozeßabschnitt den Fehlerwert für jede Einheit in der Ausgangsschicht und in der Zwischenschicht berechnet,

der Fehler δ_{oj} jeder Einheit j in der Ausgangsschicht L_O nach der Formel

$$\delta_{oj} = (t_{pj} - O_{oj})\, O_{oj}\, (1 - O_{oj})$$

berechnet wird,

der Fehler δ_{Hj} jeder Einheit der Zwischenschicht L_H nach der Formel

$$\delta_{Hj} = O_{Hj}(1 - O_{Hj})\, \Sigma_k(\delta_{Ok} \cdot W_{kj})$$

berechnet wird,

der Kopplungskoeffizient W_{ji} der Einheiten in der Zwischenschicht und in der Ausgangsschicht durch die Formel

$$W_{ji(n+1)} = W_{ji(n)} + \text{DELTA } W_{ji(n)}$$

gegeben ist,

der Lernprozeß wiederholt ausgeführt wird, bis die Gesamtsumme E des quadratischen Fehlers zwischen dem als Lehrsignal gelieferten gewünschten Ausgangswert und dem Ausgangssignal hinreichend klein wird,

dadurch gekennzeichnet,

daß der Lernprozeßabschnitt (20; 40) eine Lernvariable β_j für jeden Kopplungskoeffizienten W_{ji} jeder Einheit j in der Zwischenschicht und in der Ausgangsschicht für alle ihre Eingangswerte O_j berechnet:

$$\beta_j = 1 / (\Sigma(O_{pi}^2) + 1)$$

und daß für alle diese Einheiten in der Zwischenschicht und in der Ausgangsschicht die Variante DELTA W_{ji} des Kopplungskoeffizienten W_{ji} unter Benutzung dieser Lernvariablen β_j berechnet wird:

$$\text{DELTA } W_{ji(n)} = N \cdot \beta_j(\delta_{pj} \cdot O_{pi}) + \alpha \cdot \text{DELTA } W_{ji(n-1)}$$

worin N die Lernkonstante, α die Stabilisierungskonstante zur Verringerung der Fehleroszillationen und n die Anzahl der Lernvorgänge bedeuten.

2. Signalverarbeitungssystem nach Anspruch 1, bei dem jede der Einheiten in der Zwischenschicht und in der Ausgangsschicht außerdem weitere Kopplungen aufweist, die jeweils mit Verzögerungsmitteln ausgestattet sind, so daß ein rekurrentes Netzwerk gebildet wird mit

   einer Schleife LP, um ihren Ausgangswert O_j über die genannten Verzögerungsmittel als einen ihrer Eingangswerte zurückzuführen,

   und einem Rückkopplungspfad FB, um ihren Ausgangswert O_j als einen Eingangswert einer anderen Einheit in der gleichen Schicht oder in der Zwischenschicht zuzuführen.

3. Signalverarbeitungssystem nach Anspruch 1 oder 2, bei dem in dem Lernprozeßabschnitt (20; 40) Steuermittel zur Vergrößerung der Zahl der Einheiten der Zwischenschicht vorgesehen sind, und bei dem der

Lernprozeßabschnitt den Lernprozeß für die Kopplungskoeffizienten W_{ji} ausführt, wenn er eine Vergrößerung der Zahl der Einheiten der Zwischenschicht veranlaßt.

Revendications

1. Système de traitement de signaux, comprenant :

une section de traitement de signaux (10 ; 30) constituée d'un réseau neuronal multicouche ayant trois couches : une couche d'entrée L_I, une couche intermédiaire L_H, et une couche de sortie L_O, les couches étant constituées respectivement d'unités u_{I1} à u_{Ix}, u_{H1} à u_{Hy}, u_{O1} à u_{Oz}, chaque unité correspondant à un neurone, le réseau étant constitué de couplages dirigés vers l'avant entre les unités, chacune des unités j de la couche intermédiaire et de la couche de sortie étant conçue de façon à délivrer, pour une forme d'entrée p introduite dans la couche d'entrée, un signal de sortie O_{pj} représenté par une fonction sigmoïde selon la formule :

$$O_{pj} = 1/[1 + \exp\{-(net_j + \theta_j)\}]$$

pour la somme totale net_j des signaux d'entrée, où

θ_j est une valeur de seuil,

netj est la somme totale des signaux d'entrée fournis à une unité j de la couche intermédiaire et de la couche de sortie, et O_{pj} est le $j^{ème}$ élément de la forme de sortie réelle produite par le représentation de la forme d'entrée,

le système comprenant en outre :

une section de traitement d'apprentissage (20 ; 40) qui exécute un traitement d'apprentissage utilisant un algorithme d'apprentissage de rétropropagation, le traitement consistant à modifier les coefficients de couplage W_{ji} de toutes les unités j de la couche intermédiaire et de la couche de sortie à l'aide d'un coefficient de variation ΔW_{ji} de façon à minimiser la somme totale des erreurs quadratiques existant entre la valeur de sortie réelle O_{oj} de l'unité j de la couche L_O produite à partir d'une forme de signal d'entrée et la valeur de sortie souhaitable t_{pj} (signal d'enseignement) pour ladite unité j de la couche L_O, si bien que W_{ji} est le poids du signal de la $i^{ème}$ unité à la $j^{ème}$ unité,

la section de traitement d'apprentissage (20 ; 40) recevant une valeur de sortie voulue t_{pj}, au titre d'un signal d'enseignement, pour la valeur de sortie O_{oj} d'une unité j de la couche de sortie L_O, pour les formes p des signaux d'entrée introduites dans la couche d'entrée,

la section de traitement d'apprentissage calculant la valeur d'erreur pour chaque unité de la couche de sortie et de la couche intermédiaire,

l'erreur δ_{oj} de chaque unité j de la couche de sortie L_O étant calculée par la formule :

$$\delta_{oj} = (t_{pj} - O_{oj})O_{oj}(1 - O_{oj}),$$

l'erreur δ_{Hj} de chaque unité j de la couche intermédiaire L_H étant calculée par la formule :

$$\delta_{Hj}=O_{Hj}(1-O_{Hj})\sum_k(\delta_{Ok}.W_{kj}),$$

le coefficient de couplage W_{ji} des unités de la couche intermédiaire et de la couche de sortie étant donné par la formule :

$$W_{ji(n+1)} = W_{ji(n)} + \Delta W_{ji(n)},$$

le traitement d'apprentissage étant exécuté de façon répétée jusqu'à ce que la somme totale E des erreurs quadratiques existant entre le signal de sortie souhaité qui est fourni au titre du signal d'enseignement et le signal de sortie devienne suffisamment petite,

caractérisé en ce que :

la section de traitement d'apprentissage (20 ; 40) calcule une variable d'apprentissage β_j pour chaque coefficient de couplage W_{ji} de chaque unité j de la couche intermédiaire et de la couche de sortie pour tous ses signaux d'entrée O_j :

$$\beta_j = 1/(\Sigma(O_{pi}^2) + 1),$$

et en ce que, pour toutes ces unités de la couche intermédiaire et de la couche de sortie, le coefficient de variation ΔW_{ji} du coefficient de couplage W_{ji} est calculé à l'aide de ladite variable d'apprentissage β_j:

$$\Delta W_{ji(n)} = N.\beta_j(\delta_{pj}.O_{pi}) + \alpha.\Delta W_{ji(n-1)}$$

où N est la constante d'apprentissage, α est la constante de stabilisation permettant de réduire les oscillations des erreurs, et n est le nombre de répétitions des opérations d'apprentissage.

2. Système de traitement de signaux selon la revendication 1, où chacune des unités de la couche intermédiaire et de la couche de sortie est en outre dotée d'autres couplages ayant chacun un moyen retardateur, de manière qu'il soit formé un réseau récurrent, comportant :
   une boucle LP servant à produire, via ledit moyen retardateur, son signal de sortie O_j au titre de l'un de ses signaux d'entrée, et
   un trajet de réaction FB servant à produire son signal de sortie O_j au titre du signal d'entrée d'une autre unité de la même couche ou de la couche intermédiaire.

3. Système de traitement de signaux selon la revendication 1 ou 2, où des moyens de commande sont prévus dans ladite section de traitement d'apprentissage (20 ; 40) afin d'augmenter le nombre des unités de ladite couche intermédiaire, et où ladite section de traitement d'apprentissage effectue le traitement d'apprentissage du coefficient de l'intensité de couplage W_{ij} tandis que ladite section de traitement d'apprentissage fait en sorte que le nombre des unités de la couche intermédiaire augmente.

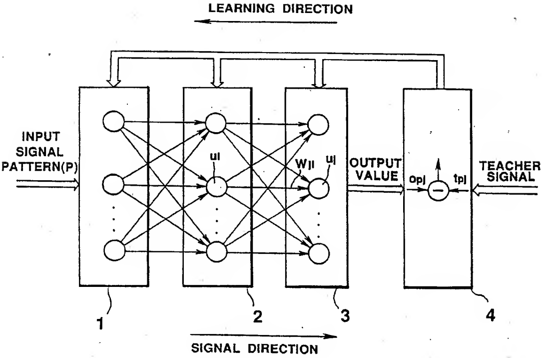

LEARNING DIRECTION
INPUT SIGNAL PATTERN(P)
ul
Wji
uj
OUTPUT VALUE
opj
tpj
TEACHER SIGNAL
1
2
3
4
SIGNAL DIRECTION

FIG. 1

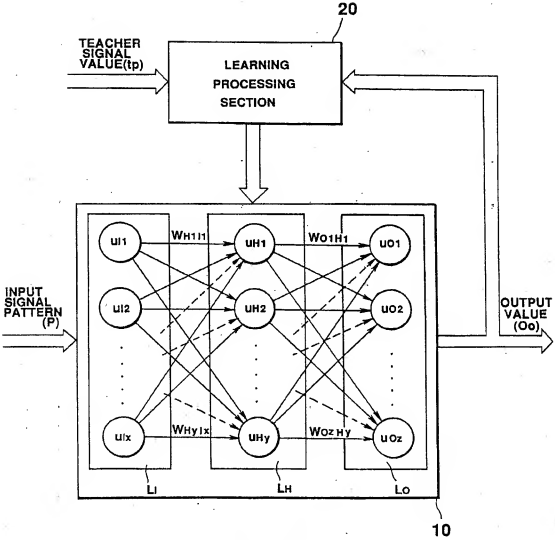

20
TEACHER SIGNAL VALUE(tp)
LEARNING PROCESSING SECTION
INPUT SIGNAL PATTERN (P)
OUTPUT VALUE (Oo)
uI1
uI2
uIx
uH1
uH2
uHy
uO1
uO2
uOz
WH1I1
WO1H1
WHyIx
WOzHy
LI
LH
LO
10

FIG.2

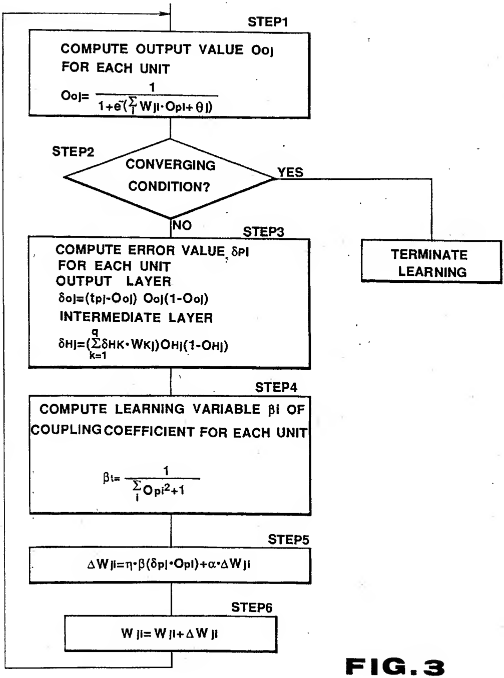
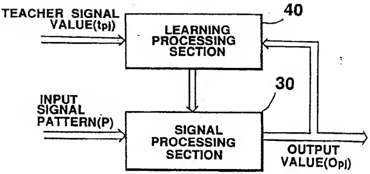
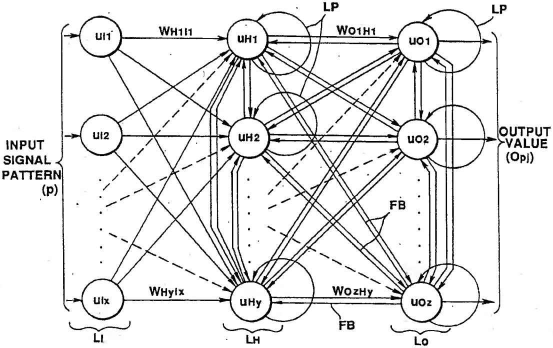

STEP1
COMPUTE OUTPUT VALUE Ooj FOR EACH UNIT
$O_{oj} = \dfrac{1}{1+e^{-(\sum_i W_{ji} \cdot O_{pi} + \theta_j)}}$
STEP2
CONVERGING CONDITION?
YES
NO
TERMINATE LEARNING
STEP3
COMPUTE ERROR VALUE δpj FOR EACH UNIT
OUTPUT LAYER
$\delta_{oj} = (t_{pj} - O_{oj})\, O_{oj}(1 - O_{oj})$
INTERMEDIATE LAYER
$\delta_{Hj} = (\sum_{k=1}^{q} \delta_{HK} \cdot W_{Kj}) O_{Hj}(1 - O_{Hj})$
STEP4
COMPUTE LEARNING VARIABLE βi OF COUPLING COEFFICIENT FOR EACH UNIT
$\beta_t = \dfrac{1}{\sum_i O_{pi}^2 + 1}$
STEP5
$\Delta W_{ji} = \eta \cdot \beta(\delta_{pj} \cdot O_{pi}) + \alpha \cdot \Delta W_{ji}$
STEP6
$W_{ji} = W_{ji} + \Delta W_{ji}$

FIG.3



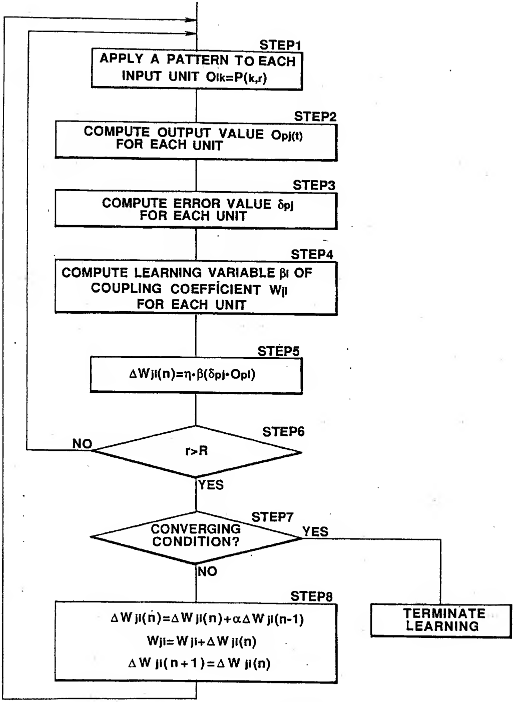
TEACHER SIGNAL VALUE(tpj)
40
LEARNING PROCESSING SECTION
30
INPUT SIGNAL PATTERN(P)
SIGNAL PROCESSING SECTION
OUTPUT VALUE(Opj)

FIG.4



LP
LP
uI1
WH1I1
uH1
WO1H1
uO1
INPUT SIGNAL PATTERN (p)
uI2
uH2
uO2
OUTPUT VALUE (Opj)
FB
uIx
WHyIx
uHy
WOzHy
uOz
FB
LI
LH
Lo

FIG.5



STEP1
APPLY A PATTERN TO EACH INPUT UNIT Oik=P(k,r)
STEP2
COMPUTE OUTPUT VALUE Opj(t) FOR EACH UNIT
STEP3
COMPUTE ERROR VALUE δpj FOR EACH UNIT
STEP4
COMPUTE LEARNING VARIABLE βi OF COUPLING COEFFICIENT Wji FOR EACH UNIT
STEP5
ΔWji(n)=η·β(δpj·Opi)
STEP6
r>R
NO
YES
STEP7
CONVERGING CONDITION?
YES
NO
STEP8
ΔWji(n)=ΔWji(n)+αΔWji(n-1)
Wji=Wji+ΔWji(n)
ΔWji(n+1)=ΔWji(n)
TERMINATE LEARNING

FIG.6

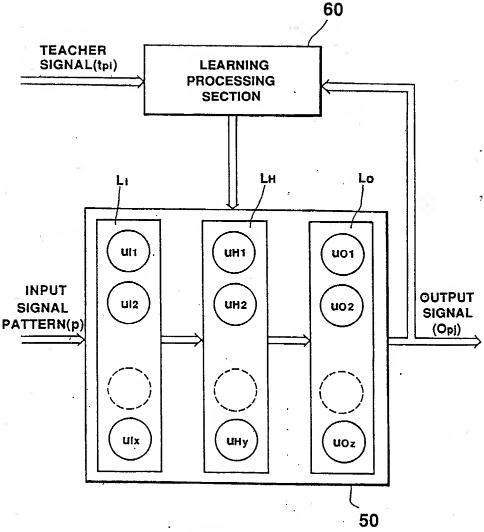

60
TEACHER
SIGNAL(tpi)
LEARNING
PROCESSING
SECTION
LI
LH
LO
uI1
uI2
uIx
uH1
uH2
uHy
uO1
uO2
uOz
INPUT
SIGNAL
PATTERN(p)
OUTPUT
SIGNAL
(Opj)
50

FIG.7

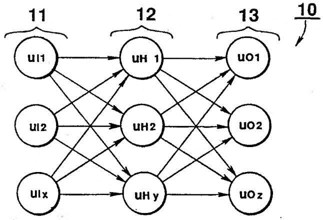

11
12
13
10
uI1
uH 1
uO1
uI2
uH2
uO2
uIx
uH y
uOz

FIG.8(A)

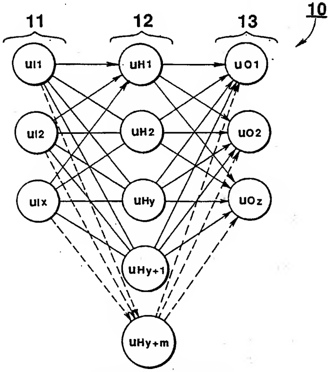

11
12
13
10
uI1
uH1
uO1
uI2
uH2
uO2
uIx
uHy
uOz
uHy+1
uHy+m

FIG.8(B)

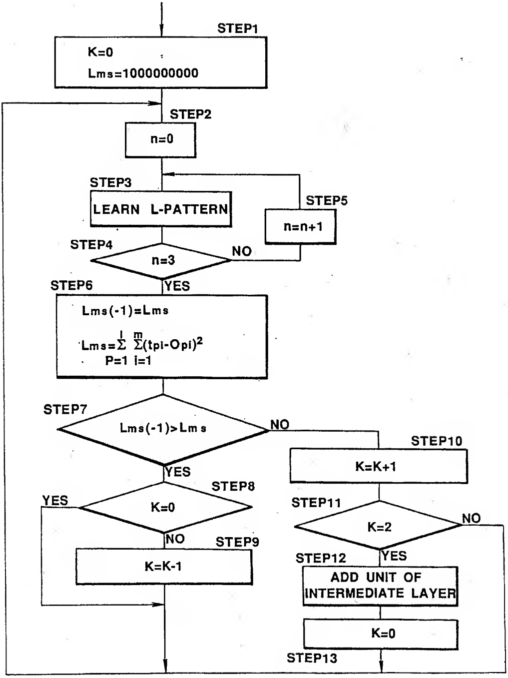

STEP1
K=0
Lms=1000000000
STEP2
n=0
STEP3
LEARN L-PATTERN
STEP5
n=n+1
STEP4
n=3
NO
YES
STEP6
Lms(-1)=Lms
Lms=Σ Σ(tpi-Opi)2
P=1 i=1
l m
STEP7
Lms(-1)>Lms
NO
YES
STEP10
K=K+1
STEP8
K=0
YES
NO
STEP9
K=K-1
STEP11
K=2
NO
YES
STEP12
ADD UNIT OF
INTERMEDIATE LAYER
K=0
STEP13

FIG.9

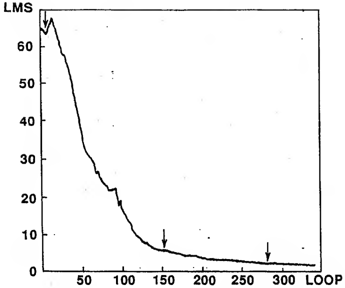

LMS
60
50
40
30
20
10
0
50
100
150
200
250
300
LOOP

FIG. 10

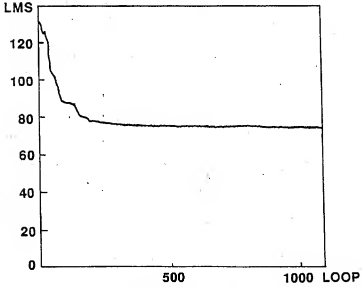

LMS
120
100
80
60
40
20
0
500
1000
LOOP

FIG.11

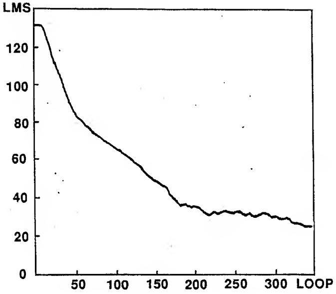


FIG.12